INDEX CHAPTER I

CHAPTER I
INTRODUCTION

A PARASITES' PAST:
PLASMODIUM VIVAX MALARIA AND HUMAN MIGRATIONS

"I recognised you as soon as I saw you
As if we called each other
The DNA in my blood is telling me
That it's you I've been looking for." – BTS, DNA

---

1. 1. INTRODUCTION

With increased ease and access to genome sequencing, there has been growing interest in understanding past human migrations through the relatively novel method of studying the genetics of human diseases (Curtin 1968, Brockmann *et al.* 2006, Tatem *et al*. 2006, Bryant *et al.* 2007, Moodley *et al.* 2009, Culleton *et al*. 2011, Bos *et al*. 2011, Miao *et al.* 2011, Comas *et al.* 2013, Taylor *et al*. 2013). Many major infectious diseases are thought to have originated around the time of the emergence of agriculture, some 10,000 years ago (ya), and to have spread with subsequent human population expansion and migration (Wolfe *et al.* 2007). Some diseases, however, are more ancient. Certain strains of tuberculosis (*Mycobacterium tuberculosis*) were present in humans before their exit out of Africa and spread with the earliest known human migrations, some 62,000-75,000 ya, as well as with the subsequent large migration around 25,000-38,000 ya (Comas *et al.* 2013). Malaria is another disease that has been strongly affiliated with humans - and probably hominids before that - for many hundreds of thousands of years (Garnham 1966 in Groube 1993). It too is now believed to have travelled with the earliest migrations out of Africa into Asia (Miao *et al.* 2011), especially with recent evidence for Africa as its place of origin (Rayner *et al.* 2011, Culleton *et al.* 2011, Prugnolle *et al*. 2013). This thesis seeks to uncover which, if any, of the major human or even hominid migrations was associated with the spread of *Plasmodium vivax* (*Pv*) malaria, by studying the phylogenetics of the *Pv* mitochondrial genome and relating it to what is known of the human past through different disciplines, including archaeology.

At the present time, malaria is topical across a broad range of disciplines. The interest surrounding it has been sparked by a marked decrease in the effectiveness of eradication efforts. For *Plasmodium falciparum* (*Pf*), the most prolific and malignant human malaria species, this is partially due to the

parasite's increasing resistance to the current drugs available (White 2004, Strode *et al*. 2014, Blasco et al. 2017, Imwong *et al.* 2017, Ochomo *et al*. 2017). For *Pv*, choloroquine resistance is also increasing but vivax malaria remains highly susceptible to other antimalarial treatments such as artemisinin combination therapy, primaquine and tafenoquine. Effectiveness of eradication efforts is caused by a combination of resistance, lack of investments that mean lack of support for poor health systems that cannot deliver the interventions effectively, and the fact that much of the remaining malaria (at least in Asia and Americas) is in geographically or demographically difficult to reach populations – such as on unstable border regions and mobile migrant populations.

In addition, certain anophelene mosquito vectors - key to malaria transmission - are developing resistance to the chemicals used in bed nets (Sokhna *et al.* 2013, Miles *et al*. 2017). This, and resistance of parasites to treatment, forms issues for regions currently infected by malaria, but also poses a threat for regions which have already eradicated the disease but remain receptive to malaria transmission. The threat to malaria-free status comes in several forms for certain regions, depending on their current status. For example, Taiwan, Hong Kong and Okinawa no longer have malaria parasites circulating in their population. However, this elimination has taken place without removing the resident transmission vectors (Yip 2009). Reintroduction of the disease from surrounding areas still occurs on low levels, but these events can as yet be easily controlled with medication. An example of new outbreaks occurring in previously malaria-free regions include a small *Pv* outbreak in the Southern Peloponnese, Greece, in 2009 (Andriopoulos 2013), and, since then, an increase in *Pv* patients across several other regions in Greece (Dimopoulou 2016), where effective *Anopheles* mosquito vectors are still prevalent. Removing effective medication from the equation has the consequence of reversing elimination as these populations become re-infected. To counteract the adaptation of the mosquito vectors and malaria parasites to the existing medication and pesticides, there has been a marked reinvigoration of the global malaria eradication programme. A global eradication programme was run by the World Health Organisation (WHO) in the 1950s and 1960s, but was abandoned in 1969 due to slow progress and the development of resistance in the anophelene mosquitoes to the pesticide DDT (Yip 2009). With support from the Global Funds for AIDS, TB, Malaria – and other bilateral donors – many countries have re-intensified their malaria control programs in the early 2000s, leading to dramatic reductions in malaria burden in numerous parts of the world (World Health Organization 2008, Global Partnership to Roll Back Malaria *et al*. 2009). This has prompted all countries in the Asia-Pacific and Americas to re-declare malaria elimination as a formal goal. For reasons stated above, unfortunately, since 2016 the progress towards malaria control and elimination has stalled and is at risk of going backwards (World Health Organization 2018).

At such a critical time, it is of great importance to understand malaria and the dynamics behind its spread and diversity. Medical research generally focuses on understanding the current situation on local scales, molecular levels and curing or preventing the disease. However, this thesis aims to help understand the long-term evolution and impact of the disease, approaching the subject from a broader perspective and much longer time scale than most. This chapter has the function of setting out where the knowledge of the past of malaria and humans in the Eastern Hemisphere currently stands: to provide a framework for further in-depth study of different forces acting on and between malaria and humans that could have influenced the genetic diversity of both species. The approach in later *Chapters 2 & 3*, using geographical mapping of malaria phylogenetics, is designed to assist in understanding the routes and timing of the spread of the disease. In addition, once the historic connection between humans and malaria has been established, the connection can possibly be used to clarify past global human migration routes in a similar fashion to previous studies of disease phylogeny (Curtin 1968, Brockmann *et al.* 2006, Tatem *et al*. 2006, Bryant *et al.* 2007, Bos *et al*. 2011, Miao *et al*. 2012, Comas *et al.* 2013).

Linking the presence of malaria to certain migrant populations also provides an opportunity to confirm the minimum extent of time over which the migrants have been exposed to the disease. This will aid in understanding the process of human genetic adaptation to a disease such as malaria, by giving an indication of the varying exposure times of the different migrant populations to the disease. An understanding of the period of exposure will, in turn, clarify to what extent malaria played a role in shaping the human cultural landscape in the region. As such, combining the datasets on humans and malaria is necessary in order to elucidate the role of disease in human evolutionary and cultural processes, and to gain insight into the role of human past activities in the spread and diversity of this disease. Before this can be done, however, it is important to first understand the timing and dynamics behind the spread and persistence of the disease throughout the globe.

Current research into *Pv* phylogenetics is starting to shed light on these interconnected pasts of humans and malaria on a global scale. Past studies with *Plasmodium falciparum (Pf)* have already indicated an introduction of the disease to the Americas with the African slave trade (Yalcindag *et al*. 2012). For *Pv*, the African-American connection in the context of human migration was recently published by Rodrigues *et al*. (2018). A possible European origin for American *Pv* was discussed in Gelabert *et al.* (2016), with the sequencing of a historic sample of *Pv* from Spain. Even so the origin of South American *Pv* and how it connects to *Pv* in the rest of the world remains unclear, possibly due to the lack of European samples. The South American connections are studied extensively by others (Taylor *et al.* 2012, Rodrigues *et al.* 2018) and the complex positioning of American data in the *Pv* mtDNA phylogeny as observed in other studies (Mu *et al*. 2005, Culleton *et al*. 2011, Miao *et al*. 2012) may obscure some of the finer details found in other

regions. This thesis concentrates exclusively on *Pv* in the Eastern Hemisphere: the African, Eurasian and Oceanian continents, from the Atlantic to the Pacific.

Possibly more substantial evidence for a connection between human migration and *Pv* distribution is the presence of a divide in Central, Eastern and South East Asian *Pv* versus an Indo-Malay, Australasian branch of *Pv* (Miao *et al.* 2012). The divide could be the result of two main human migration events out of Africa and into Asia following different routes at different times, as has been found with tuberculosis (Comas *et al.* 2013) and indicated by human whole-genome variation (Gignoux *et al*. 2011, Stewart & Chinnery 2015) and archaeology (Bae *et al.* 2017).

A recent paper published by Skoglund & Mathieson (2018) reviews the evidence for human migratory movements in the context of currently available publications on ancient DNA in hominids. The review reiterates the origin of modern humans from Africa, with all modern humans outside of Africa falling within one genetic clade that branches off a diverse African phylogeny. Though on a single branch off the African tree, the population movement and diversification of the Eurasian (and American) clade has been varied and complex. It has included an early split that has shaped the variety of populations on the very eastern edges of the Eurasian and Oceanian continents. This split saw two major population movements: one taking a southern route along the Indian Ocean coast towards and into Australasia, also known as haplogroup M; the other took a more northerly route, probably through the Levant and possibly up through central Asia and along the northern side of the Himalayas, also known as human haplogroup N (Stewart & Chinnery 2015, Bae et al. 2017). A third haplogroup R emerged from haplogroup N and gives rise to several lineages in Europe and eastern Asia whilst it takes a similar northern route to haplogroup N (Stewart & Chinnery 2015).

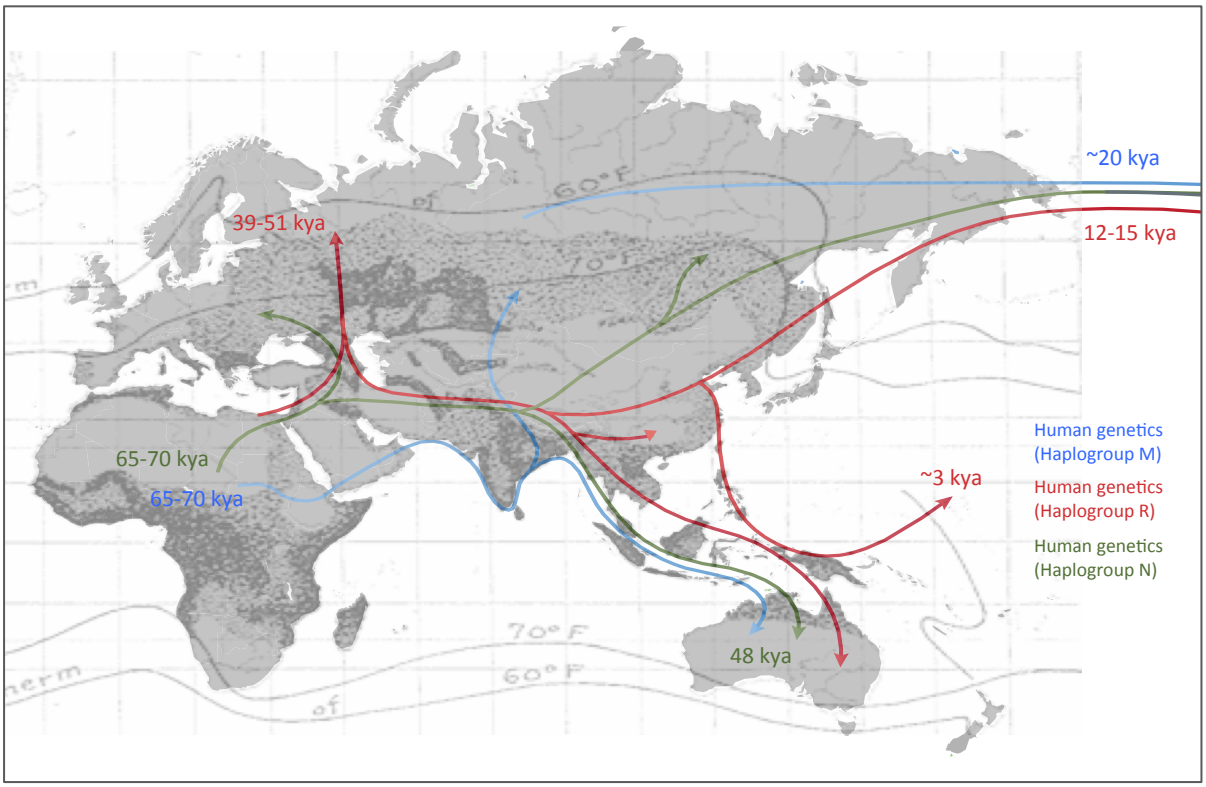

Figure 1: Broad indication of the migratory routes and timings of the earliest *Homo sapiens* to exit Africa, based on Bae *et al.* 2017. Routes as found for the different human mitochondrial genome haplogroups are given in light blue (haplogroup M), red (haplogroup R), and green (haplogroup N). Dates are estimated ages of time to the most recent common ancestor of populations in the region marked. Overlain with the malaria distribution map by Boyd 1930, indicating the presence of both human infecting malaria parasites as well as anophelene vectors. Shaded/dotted regions denote areas with malaria infections: increased density indicates higher number of infections across the human population in the region.

Assuming that malaria vectors were present along both routes, as they were until recently (Boyd 1930, MacDonald 1957, Sinka *et al.* 2012), the expectation is that if *Pv* came out of Africa with all migrants, it had the opportunity to travel via both migration routes. This can be seen in *Figure 1*, where most estimated migration routes pass through regions that are, or until recently were, malarious. Climate throughout the past 100 ky was changeable (Timmerman & Friedrich 2016, Stewart *et al.* 2017), so consistency of vector availability is unpredictable. Yet at times more suitable habitat for transmission existed in the tropical and subtropical regions; there is much evidence for deserts having been less extensive than they are today, e.g. at 95 - 107 and 75 - 90 kya (Timmermann & Friedrich 2016) and during the Neolithic, 12 - 6 kya (Platt *et al.* 2017). Deserts appear to be a huge barrier for malaria survival, as seen in *Figure 1*, with absence of malaria in the Sahara in Northern Africa, the central areas of the Arabian Peninsula and the Gobi desert, in Southern Mongolia and Northeastern China. The spread of *Pv* across the Eastern Hemisphere is explored in greater depth in *Chapter 2*, which looks at the global distribution of *Pv* according to its mitochondrial genome as influenced by several factors, including human migration.

From the two separate global human migration events, the first wave of *Homo sapiens* out of Africa connects to the first migration into Australasia by anatomically modern humans. Few traces have

remained in intermediate geographic regions through a high density of subsequent migrations. If vivax malaria distribution patterns mirror that of *H. sapiens*, the closest genetic relations of *Pv* from the Australasian region are expected to be found at its original source, in Africa. Both malaria and human populations across the region display especially high diversity around Melanesia, the region in the Pacific encompassing the islands of Papua, the Solomon Islands, Vanuatu and Fiji. Much research into human diversity has focused on the varying origins of the human populations in Melanesia (Bellwood *et al.*1995, Spriggs 1997, Terrell *et al*. 2001, Pawley *et al.* 2005, Kirch 2010, Donohue & Denham 2010, Gignoux *et al*. 2011, Spriggs 2011, Xu *et al*. 2012). These varying origins represent the main determinant of the large-scale diversity in the Indo-Pacific region, both in cultural and in genetic terms.

Many areas of research have looked into the migratory movements and timings of people into the Indo-Pacific and Melanesia, as there is much focus and ongoing interest in the timing of migrations 'Out of Africa'. Archaeological, geological, palaeoecological, climatological, linguistic and genetic studies including modern human, ancient, commensal/agricultural and, recently, disease genetics, have all contributed to what is currently known about past human migratory movements (Jin & Su 2000, Larson *et al*. 2007, Atkinson *et al*. 2008, Dobney *et al*. 2008, Petraglia *et al.* 2010, Bellwood 2011, Bellwood *et al.* 2011, Gignoux *et al*. 2011, Rasmussen *et al*. 2011, Henn *et al*. 2012, Jinam *et al*. 2012, Moodley *et al*. 2012, Paraskevis *et al*. 2012, Soares *et al*. 2012, Stewart & Stringer 2012, Comas *et al*. 2013, Malaspinas *et al*. 2016, Skoglund *et al*. 2016, Soares *et al*. 2016, Gomes *et al.* 2017). It has been recognised that at least two main migratory movements of anatomically modern humans occurred into Oceania (e.g. Chappell 1993, Spriggs 1997, Kayser *et al*. 2006, Hudjashov *et al*. 2007, Duggan *et al*. 2014, Gomes *et al*. 2015). 'Oceania' is used here in a broad sense, referring to all the tropical island groups between Asia and the Americas, yet pertaining especially to the Melanesian region.

These two migratory waves into Oceania connect back to the two different global migration routes, and the migrations are represented by a duality in linguistics and human genetics. The two-way split is shown in the Australoid/Papuan versus the Austronesian populations. The Australoids first entered the region 75,000 to 40,000 ya (Kirch 2010, Summerhayes *et al*. 2010, Gomes *et al*. 2015, Clarkson *et al*. 2017, Larruga *et al*. 2017, Nagle *et al*. 2017) with possible subsequent migrant waves between 15,000 and 5,000 ya, represented by - intraregional movement of - Papuans (Soares *et al*. 2008, Gomes *et al*. 2015) who did not reach Australia as sea-levels between Australia and Papua New Guinea had heightened. The much later Austronesian migration came to the region from 4,000 ya, moving far out into the Pacific and probably towards Madagascar across the Indian Ocean (Serjeantson *et al.* 1982, 1983, 1992, Serjeantson 1984, Kirch 2010). The Austronesian migration into Australasia and specifically Melanesia at 4,000 ya is believed to have strong connections with the *H. sapiens* that took the northern route and settled in East Asia, though its exact origins are under debate (Spriggs 1993, Bellwood *et al.*1995, Terrell *et al*. 2001, Kirch 2010, Donohue & Denham 2011, Spriggs 2011, Xu *et al*. 2012, Stewart & Chinnery 2015, Bae *et al*. 2017).

Apart from this larger scale variation, however, there is a conspicuous representation of smaller scale differentiation between population groups in Melanesia. This differentiation is present in languages, cultural practices and genetics (Serjeantson *et al.* 1992, Spriggs *et al*. 1993, Spriggs 1997, Pawley *et al.* 2005). In this region, however, besides general genetic variations, populations display many different genetic adaptations to coping with malaria (Serjeantson *et al*. 1992, Howes *et al*. 2013). Malaria has played a significant selective role in the evolution of human genetic blood defects, which act as an adaptive mechanism against the more adverse consequences of malarial infection (Flint *et al*. 1986, Flint *et al.* 1998, Kwiatkowski 2005, Williams 2006). The recent discovery of a genetic polymorphism found only in Austronesian people that is associated with a decrease in the risk of contracting vivax malaria (Rosanas-Urgell *et al.* 2012) suggests that the past relationship between human populations and malaria across Austronesia merits further study. As such, *Chapter 3* of this thesis will focus on discovering when and with which human migrations *Pv* entered the Melanesian region.

So far, it has been shown that establishing when malaria entered regions across the globe is important to developing an understanding of its role in shaping human populations in those regions. Doing so can help explain some of the larger-scale diversity found in malaria in the region, as well as creating an opportunity to gain insight into long-term interactions between human populations and disease. As has been mentioned previously, malaria is a disease that significantly influences human populations and exerts strong selective pressure on the human genome to adapt and survive the consequences of the disease. In addition, it would be expected that malaria could select favourably for certain behavioural adaptations. It follows that the earlier malaria was present in human populations, the longer humans would have had to adapt to it, including both the rise of genetic mutations which lessen the severity of the disease, and behavioural changes which disrupt transmission. Viewing society from the perspective of long-term malarial selection could help us understand not only human -cultural- evolution, but also give insight into how societies might operate at their most efficient whilst dealing with the burden of malarial disease.

1. 2. DNA RESEARCH IN ARCHAEOLOGY

As Jobling (2012) explains, there are two methods of using information from genomes to interpret the past: (1) by studying ancient DNA directly; or (2) by mathematically reconstructing the patterns revealed in modern DNA. In archaeology, any material studied is usually at least a few centuries old. The first way to read information in DNA is thus relatively simple to understand in an archaeological context, as it concerns studying ancient biological material itself. It is easy to see how studying ancient DNA (aDNA) from bones, plants or other organic matter found in archaeological digs fits into the discipline of studying ancient materials to further an understanding of the past of human civilizations. DNA from hominid remains can reveal the relations between the hominid and modern humans.

Ancient genomes from anatomically modern humans found in several sites across Asia confirm some human history as deduced from modern genomes and archaeological finds, but disagree with other previously constructed dynamics. For example, the sequenced draft genome for Neanderthals (Green *et al.* 2010) provided relatively clear data indicating that around 1 to 4% of modern human Eurasian genes are likely to have come from Neanderthals, revealing possible interbreeding of Neanderthals with our *Homo sapiens* ancestors at some point in the past. This was unexpected; the current findings are that the genetic contribution of Neanderthals to human lineages outside of Africa is around 2%, and there is a 3% contribution of the still mysterious Denisovan lineage(s) in some of Asia's human populations, specifically in Oceania (Skoglund & Mathieson 2018).

What makes human ancient DNA (aDNA) especially useful is that it gives both a place and a time point to refer to in regards human movement and evolution. With aDNA, the estimated mutation rate over time for the genome gets calibration points from both time and number of mutations, making the molecular clock much more precise. In reference to information about the past obtained from human aDNA, ancient mitochondrial genomes confirm that the most recent common ancestors (TMRCAs) of African and non-African lineages can be dated back to 62-95 thousand years ago (kya), and that non-African sub-lineages diversify around 45-55 kya. These dates correspond with earlier estimates of these timings. However, some archaeological findings of anatomically modern human remains in Australia and Indonesia push back these dates considerably (Clarkson *et al.* 2017, Westaway *et al.* 2017): the archaeological finds in the Arnhemland escarpment (Australia) and Lida Ajer Cave (Sumatra, Indonesia) are dated to 65 (59.3-70.7) kya and 68 kya respectively, considerably predating the sub lineage diversification estimate above. It should be kept in mind, however, that biological processes--and especially those that deal with population dynamics, such as population genetics--are not straightforward. Genetic relationships over time are influenced by multiple variables. This usually leads

to a genetic time overestimation of a physical geographic split. This is due to a lack of representation of all data at all time points, such that the odds of finding the exact ancestor of two lineages are rather low; the representative ancestor will be found slightly farther back in time.

Similarly, when studying human disease in the past, the most straightforward method is to use ancient DNA of the disease, found either in direct association with certain hominid sites or with certain events in the past. Examples of such research include the studies tracing the origins of the Black Death, *Yersinia pestis*, by Bos *et al*. (2011) and tracing human migratory and trade movements through bowel parasites, including *Caenorhabditis elegans*, preserved in archaeological materials (Patrik Flammer, *pers. comm.* 2012). Such an approach might be possible through analysing preserved soft tissue of human remains for malaria sequences, with varying origins. Alternatively, old samples of European vivax could aid in understanding the origins of South American vivax, which could have been introduced from Spain or Portugal instead of Africa during the age of exploration and slavery (Gelabert *et al*. 2016, Rodrigues *et al*. 2018).

Ancient or old samples are rare, but do exist. However, when such samples are available, there are further complications; when working with ancient DNA, there are often only very few samples, and the DNA tends to be severely fragmented. The information retrieved is thus both scattered and scant. Though ancient DNA helps increase the precision of time and route estimates, aiding significantly to increasing evolutionary insight in a species, it is important not to forget that even these reconstructions rely on models and assumptions about population dynamics and are not infallible. This is where the second method of reading DNA comes into play, which relies more heavily on reconstructive models but has alternative strengths. Instead of using aDNA, insight is gained through reading several samples of modern DNA and studying the patterns that are revealed. In a way, DNA is an archaeological record on its own; it accumulates mutations over time. As chances of the same mutation occurring more than once are negligible if not under selection, it is possible to trace relatedness by comparing levels of similarity between different individuals. The problem in this method lies not in the number of samples – though this can be an extra complication due to restrictions in time and funding – but in having to infer through statistics how samples are related and what this tells us about the past. These 'phylogenetic' methods of reconstructing genetic relationships based on the DNA code – whether ancient or modern – provide an estimation of the linear associations between genetic samples across time. In both cases time is estimated: for aDNA through archaeological methods of estimating a sample's age; for modern DNA based on estimations of mutation rates (how much time passes between the accumulation of one mutation and the next).

To understand processes in the past which have led to the patterns seen across geographic space, extra dimensions can be added that help clarify time outside the information contained in the DNA. 'Phylogeography' overlays a geographical map with phylogenetic data to demonstrate the genetic relatedness between samples from their sites. Other dimensions come from data provided by archaeology, linguistics, palaeo-anthropology and the earth sciences. As Oppenheimer (2012) states: '*A combined approach should give increased confidence by triangulation*'. The combination of different classes of data help fill gaps, and provide checks through expected overlap. Data from other disciplines provide much of the broader understanding of human activities within the observed timeframe. One must, however, remain wary that information provided by one area in an integrated data-set does not form the basis for conclusions of another, as this will lead to false positives through data reiteration. Another hazardous strategy is to randomly pick the best fitting story about the past to explain the patterns in the data. It is advisable to remain aware of alternative explanations, which could be much more complex, and keep a certain amount of objectivity. Approaching a genetic dataset from a combined evolutionary and archaeological perspective, bringing an awareness of alternative explanations for genetic variation on at least two different levels prior to looking at the data, is arguably a great advantage.

1. 3. MALARIA, ORIGIN AND SPREAD

Malaria is a disease caused through infection by one or more specific species of single celled protozoan organisms, of the genus *Plasmodium*. Infections are passed to humans through anophelene mosquito vectors, *Anopheles* spp.. Both humans and mosquitoes form crucial parts to the parasites' life cycle. Sexual replication takes place in the mosquito vector. 10-18 days after infection of the mosquito, the parasites are found in 'sporozoite' form, in its salivary glands. The mosquito passes on these sporozoites to humans when feeding on them for a blood meal. In the human host, the parasites first invade and grow in the liver cells. In *Pv*, this phase in the liver can be prolonged by the parasites going into a dormant 'hypnozoite' form, for several months up to four years. From the liver cells, the parasites are released into the blood stream, where they invade and multiply in the red blood cells. Another difference between *Pv* and *Pf* is found at this stage: whereas *Pv* invades only young red blood cells (i.e. erythroplasts and reticulocytes), *Pf* can invade blood cells of different ages (Miller *et al*. 2002). After multiplying, the parasites burst out of the blood cell walls and release 'merozoite' daughter cells into the bloodstream, which reinvade and multiply in red blood cells in a repetitive cycle. This stage is what causes the symptoms associated with malaria. Another form of blood stage parasite – the 'gametocyte' – infects the mosquito when it feeds off a malaria infected person, and starts the whole cycle over again.

Malaria and human ancestors have been associated with one another from before the emergence of the hominids (Garnham 1966 in Groube 1993, Silva *et al.* 2011). Garnham (1966) contributes that the current complex equilibrium in the relationship between malaria and its definitive and intermediate hosts, respectively, has taken a long time to establish, indicating a very long period of evolution. What is currently known of malaria's past depends on the species. For many years, research has mainly focused on the most malignant malaria parasite – *Pf,* which accounts for most infections deaths worldwide and can be cultures *in-vitro*) – providing ample data to understand its past. The recent focus of research on *Pv* should aid in rapidly enlightening its origins. A brief overview of what is known for both is given below.

1. 3.1 PLASMODIUM VIVAX & PLASMODIUM FALCIPARUM: AGE & EVOLUTION

Previous age estimates calculate the origins of *Pf* to stem from before the human-chimpanzee split, infecting human ancestors in Africa for the first time between 112,000 and 1,036,000 years ago (Escalante & Ayala 1994, Joy *et al*. 2003, 2006, Silva *et al*. 2011, Baron *et al*. 2011). A species of *Plasmodium* infecting gorillas has been found to be the immediate precursor, or at least closest relative, of *Pf* (Liu *et al.* 2010, Rayner *et al.* 2011), re-igniting the discussion on apes being a recurring source of re-infection for humans by plasmodia. There have been studies concluding that *Pf* originated from primates relatively recently –

about 10,000 ya. However, the date estimates from these studies have later been attributed to the genetic effects of a relatively small, bottle-necked – but already established – *Pf* population, suddenly expanding during the onset of agriculture around 10,000 ya (*for discussions see*: Tishkoff *et al.* 2001, Volkman *et al*. 2001, Coluzzi *et al.* 2002, Barry *et al.* 2003, Joy *et al.* 2003, Hartl 2004, Baron *et al*. 2011).

Recent publications have brought to light the possibility that *Pf* travelled along with at least one of the several human migration waves into Oceania in the distant past, possibly even dating back to the earliest known migrations out of Africa between 130,000 and 40,000 ya (Joy *et al.* 2006). Conflicting ideas about when *Pf* branched out of Africa still remain. The earlier mentioned increase of *Pf* due to agricultural development around 10,000 ya, in conjunction with increased human migration at the time due to population increase, is a strong argument for a later introduction of *Pf* (Livingstone 1958, Tishkoff *et al.* 2001, Coluzzi *et al.* 2002, Hartl *et al.* 2002, Hartl 2004, Joy *et al.* 2003, Rich *et al.* 2009). Vedic texts from India (4,000 – 3,500 ya) indicate that *Pv* and *P. malariae* reached India before *Pf* (Carter & Mendis 2002). The latter is mentioned in texts around 3,000 years old.

Unlike *Pf*, the place of origin for *Pv* has been subject to much debate. Supporting evidence was present for both Asian and African heritages. High genetic diversity and the presence of closely related malaria species in primate/ape species, namely macaques, are both strong arguments to indicate an area of origin. As such, previously, evidence was slightly stronger for an Asian origin (Escalante *et al.* 2005, Mu *et al*. 2005), until the discovery of multiple forms of malaria – closely related to vivax – infecting apes in Africa (Culleton *et al*. 2011, Prugnolle *et al.* 2013, Liu *et al*. 2014). The high diversity of these forms of malaria indicates a strong possibility for an extensive African origin for *Pv*. Besides these multiple vivax-related malaria species in Africa, there is evidence for a protracted presence of vivax malaria in Africa through the presence of human genetic adaptations which specifically provide protection against vivax malaria – indicating vivax origins are expected to lie in Africa. It is also estimated to be older than first thought, largely pre-dating the first known human migrations out of Africa (Mu *et al.* 2005, Silva *et al.* 2011).

In addition to the discovery of the new vivax-like primate malaria species, the advent of molecular diagnostics has found *Pv* to be more common in western and central Africa than previously hypothesised (Culleton & Ferreira 2012, Howes *et al*. 2015, Poirier *et al*. 2016). The occurrence of *Pv* across these regions was doubted for several decades due to the presence of Duffy negativity – the specific genetic adaptation against vivax malaria, which modifies part of surface of red blood cells, making it harder for the parasites to enter – in most human populations in the area. Evidence has accumulated for *Pv* infecting people who are Duffy antigen negative (personal observation & Niangaly et al. (unpublished, 2017), Ryan *et al*. 2006, Menard *et al*. 2010, Mendes *et al*. 2011, Wurtz *et al*. 2011), which is a genetic adaptation in mainly central and western African human populations that protects against *Pv* infection. Together with the evidence

provided by the new vivax-like malaria and the widespread presence of Duffy negativity in human populations, the evidence mounts for the root of *Pv* to lie in Africa.

Besides the uncertainty of an African or Asian origin, some arguments were presented for *P. vivax* originating in the Americas. The arguments were mainly based on the genetic proximity of a *Pv*-like *Plasmodium* species, found in monkeys in the Americas. Until the recent publications by Culleton *et al.* (2011) and Prugnolle *et al.* (2013), on African vivax-type parasites in wild apes, the closest relatives of *Pv* were believed to occur in Old World macaques (*P. cynomolgi, fragile* and *knowlesi*) (Waters *et al.* 1991, 1993, Escalante & Ayala 1994, Mu *et al.* 2005), and South American monkeys (*P. simium*) (Escalante *et al.* 2005, Mu *et al.* 2005). The monkey-infecting species *P. simium* in the New World is so closely related to *Pv* that it has likely descended from the human strain, but the timing of this evolution from human to monkey is unknown. It has been implied that the introduction of *Pv* that led to the evolution of *P. simium* was pre-Columbian, and possibly connected to the Austronesian expansion through Polynesia (Escalante *et al.* 2005, Cormier 2011). However, these findings are still questionable and evidence is mounting for the introduction of *Pv* into the Americas through the European colonists (Rodrigues *et al.* 2018, Gelabert *et al.* 2016).

Over the past few years, there has been much discussion about the different ages for *Pv* and *Pf* (Ricklefs & Outlaw 2010, Silva *et al.* 2011). One argument that has been used in the past to argue a greater age for *Pv* than *Pf* is its reduced malignancy (Brooks & McLennan 1992). When a disease is adapting to a new host, it can be unusually malignant, or virulent, as it still needs to adapt to the host and the host to the disease. Killing off a host quickly, before infecting new hosts, is a strategy that hampers survival of the disease and thus generally uncommon in diseases that have evolved in conjunction with their host for a long period of time (Wolfe *et al.* 2007). However, estimated ages of both *Pv* and *Pf* date back to hominids, which puts both parasites well beyond an initial adaptive period. It is unlikely that the level of malignancy for either reflects their respective ages, instead pointing towards a difference in evolutionary strategy.

Reduction in virulence – how harmful a disease is to the host – is only one part of a suite of evolutionary strategies that usually caters towards transmission or survival of the disease. Disease transmission strategies tend to trade-off between virulence and host survival. Fast and high transmission rates are often associated with impacting the host system & behaviour in ways that increase transmission rates, often at high cost to the hosts – also known as increased virulence of the disease. Such strategies include increasing fluid or airborne transmission, with physical effects to the host that significantly increase such transmission rates including the forced release of bodily fluids (e.g. snot, vomiting, sweating, diarrhoea, or even opening up of skin and release of body fluids through open wounds/pustules), or the release of infections into the air (e.g. through coughing and sneezing).

High transmission strategies through increased virulence impacts the host's body significantly, often leading to reduced survival rates. In addition, a high transmission strategy means available hosts in the vicinity will be rapidly exposed to the disease. The combination of these has the effect of rapidly depleting the available hosts in a host population, removing infectible hosts either through survival – with assumed developed immunity – or through death. A modern complication for the survival of a disease that shows severe symptoms on infection, is the possibility for immediate medical attention. Patients can get treated, isolated or quarantined early on, reducing infection opportunities. Such isolation strategies work for *Pf*, however, *Pv* can transmit before clinical symptoms appear in infected patients, and the *Pv* cycle in humans includes a possible latent stage in the form of the hypnozoite. The hypnozoite hides in liver cells for periods of time that can last up to several years and cause continuous resurgence of the disease in a patient. There is currently only one drug that can eliminate these cells, primaquine – also known as tafenoquine (Campo et al. 2015). The issue is that many populations across infectious sites in Southeast Asia and the Pacific carry a gene – glucose-6-phosphate dehydrogenase (G-6-PD), ironically an anti-malarial genetic adaptation – that makes them more susceptible to oxidative damage to haemoglobin, and erythrocyte (young red blood cell) destruction. Leading to illness and possible death.

The alternative strategy to high transmission strategies, at the other end of the spectrum, is to focus on long-term persistence within a host. This strategy focuses on high host survival rates and long infection duration – increasing the chances of infection over time, often utilising low detection rates for an increased opportunity for exposure to a much larger host population, as hosts with low to no symptoms have relatively high mobility and could be in contact with a large selection of different populations over time. Trade-offs include the possibility of an infected host not infecting others before recovering, receiving treatment or death, forming a dead-end to the disease.

Countering the argument for a younger age due to higher virulence, these theories of evolutionary adaptation form an alternative explanation: instead of the higher virulence being a maladaptation of a young human disease still adapting to its new hosts, it is much more likely *Pf* evolved this strategy. Characteristics exhibited by *Pf* that classify it as such include its relatively short and intense virulence/pathology, indicating dependence on fast transmission and the availability of ample numbers of both hosts and vectors (Wolfe *et al.* 2007). In doing so, it is characteristic of a crowd disease which would struggle to survive in small migratory human populations. The assumption is that, with an increasing human population, *Pf* has evolved towards being an opportunistic crowd-disease, the traits of which it has managed to exploit after the onset of agriculture and the growth of human populations. This high virulence, crowd disease type strategy is also suspected to have been adopted by more modern

strains of tuberculosis (Brooks & McLennan 1992, Wolfe *et al.* 2007, Joy *et al*. 2012, Comas *et al.* 2013), contrasting with the lower transmission strategies used by older strains. Possibly indicating that some of these ancient diseases continue to adapt to novel population dynamics. *Pf* differs from the usual crowd-diseases, and can be sustained by relatively smaller populations, for two main reasons: 1) immunity to malaria requires many exposures and is only retained with constant re-infection; and 2) the mosquito vector prolongs its persistence as it complicates control over, and predictability of, infections.

The two host system in malaria complicates the already complex system of host-parasite or host-disease strategies. In malaria, the necessity for two hosts overcomes some of the issues of both high and low virulence strategies. For example, in both *Pv* and *Pf* the mosquito vector aids in sustaining transmission even if patients are immobilised. However, having to function and survive in two vastly different host species, also increases issues of having to adapt to, survive in, and control infection pathways through two very different types of species. Though the life stage forms of the malaria parasites presenting themselves inside the two hosts are different – providing opportunities for the parasite to specialise and adapt to each host in generally well-separated systems – it is important to keep in mind that some effects of malaria we see in humans might be adaptations to evolutionary pathways in mosquitoes, and vice versa.

1. 3.2 ORIGIN & THE DUFFY NEGATIVITY CONUNDRUM

As established, there is increasing evidence for Africa as the place of *Pv's* origin. Classical population genetic theory first supported an Asian origin, because generally, a species' place of origin is likely the one with the highest genetic diversity within the species itself. It is important to address how there are certain exceptions to this general, population genetics-based, assumption. Usually only a limited number of migrants from the original founding population settle new areas, causing genetic variation in migrant populations to be generally lower (Novak 2007). This decrease in genetic variation through selective migration – termed as founder effects or population bottlenecks – across a population can be mitigated in several ways. One is when multiple introductions of varying strains occur, especially if migrants from different populations accumulate in one area, or an uncommonly variable selection of the founding population is included amongst the migrants (Novak 2007, Lavergne & Molofsky 2007). Other processes influencing diversity are: 1) increased selection pressures in different directions (disruptive selection) or, 2) isolation of many smaller populations (genetic drift, and allopatric speciation through fragmentation of populations). These mechanisms could have contributed to the diversity found in *Pv* today.

There are two main identifiable factors influencing disruptive selection. Firstly, and generally the most obvious one – as it influences selection in many species – is the environment. The many different environments in which malaria is found across the globe would have forced malaria to adapt (Vellend *et al*. 2007). Not only because malaria was adapting to the environment itself, but malaria would have also been adapting to changes in behaviour, biology and availability of its vectors.

Another, more uncommon, form of disruptive selection acting on both *Pv* and *Pf* malaria species is the huge variation in species found in its mosquito hosts – as present across the globe. Disruptive selection could have been influenced by the many different species of mosquito present in different regions of the world that can act as vectors for malaria, where malaria strains adapt more towards one vector than another (Joy *et al*. 2008). The different environments found in the island Southeast Asian (ISEA) and Melanesian areas would have had a similar effect, impacting host migration as well as vector distribution (Vellend et al. 2007). Different mosquito species and the environment could also both have been drivers of allopatric speciation, i.e. isolating populations through distribution across different mosquito species or island geographies, mountain ranges and varying environments that limit human or vector migration.

If vivax malaria was introduced during the first human migrations, there would have been time for Asian *Pv* mtDNA to diversify beyond the original sequences introduced by the colonising population, especially when combined with the processes mentioned above. Besides the influence of the previously discussed processes of adaptation and diversification of the parasite, on the distribution patterns of *Pv* mtDNA seen today, it is possible that dynamics in the vectors contributed to its distribution too. The African *Pv* founder population could have been burdened with negative selection pressures and have declined to smaller numbers in the past. The strongest suggested cause of this negative selection pressure is that humans in the area developed Duffy negativity. Duffy negativity is a genetic mutation in a critical receptor for *Pv* invasion, and conferring very strong – albeit only partial – protection against vivax malaria. The occurrence of a high prevalence of Duffy negativity across especially Western and Central Africa is taken as evidence for an extended joint past of vivax malaria and humans in the region (Livingstone 1984), with Duffy negativity a human adaptation to high levels of exposure to, and selection from, *Pv*.

As Duffy negativity prevents the malaria parasites from effectively invading the hosts' red blood cells, the absence of Duffy basically bars the parasite from replicating. This not only prevents the host from getting ill, but it also prevents the parasite from invading mosquitoes and moving on to a new host, thus preventing the spread of malaria. The strong protection against vivax malaria that Duffy negativity confers to its human carriers means that a growing Duffy negative population of human hosts complicates the spread and survival of vivax malaria. An increasing number of Duffy negative human

individuals within a population confers what is known as 'herd immunity', protecting the population from infection. Chances of an increasingly rare Duffy positive host getting infected by a random, infected mosquito, become significantly reduced. Even were this infection to occur, the chances of infection occurring in sequence – which is required for the survival of the disease in a population – would also be very much reduced.

Even if some (low) levels of infection do occur in the Duffy negative human host, as has been increasingly observed (personal observation, Ryan *et al*. 2006, Menard *et al*. 2010, Mendes *et al*. 2011, Wurtz *et al*. 2011, Albsheer *et al*. 2019) to transfer the malaria parasite to mosquitoes a certain density of parasites is required. High parasite densities are also required further downstream, for mosquitoes to pass on the infection effectively back to humans. As such, a significant decrease in parasite density within an infected host complicates the spread of the parasite, and thus malarial disease; besides protecting the infected host itself from the negative effects of malarial disease. Duffy negativity, therefore, is believed to have all but disappeared – and at the very least gone through a severe population bottle-neck, which lowered genetic diversity – in Western and Central African populations, at some point in the past.

Since the discovery of Duffy negativity in the 1970s (Miller *et al*. 1976), it has been taken for granted that vivax malaria was hardly present in those areas of West Africa – and many other parts of Africa – with high levels of Duffy negativity. The underlying issue was the assumption that *Pv* would be unable to sustain populations in areas containing high levels of resistance in humans. However, recent publications have shown that *Pv* can still infect Duffy negative individuals (personal observation, Ryan *et al*. 2006, Menard *et al*. 2010, Mendes *et al*. 2011, Wurtz *et al*. 2011), and that apes might have acted as possible reservoirs for the disease in the distant past (Prugnolle *et al*. 2013) – indicating the possibility of *Pv* presence even in Duffy-negative areas of Africa being sustained through time. Whether the reduced, but still present, invasive properties of *Pv* within Duffy negative individuals is a new adaptation or one which has been around for a long time, is unclear. However, Boyd (1930) shows that vivax malaria was still present in large areas of Africa, including Duffy negative regions, around the start of the 20th century. It is, however, notable that at that time, confusion of *Pv* with *P. ovale* was rife. Even so, there is value in the reflection that observations by Boyd and earlier malarialogists were (dis)missed in favour of novel popular theories, demonstrating that certain scientific theories can find traction beyond the observed, even with the plausible presence of counter-evidence. Even so, extensive observations, going back several decades, exist of Duffy-positive travellers from Europe acquiring *P. vivax* infections during their visits to West or Central Africa (e.g. Baranova *et al*. 2019), indicating that *P. vivax* has been persisting at low levels throughout Africa for a prolonged period of time. Irrespective of mode of persistence of *P. vivax* in Africa, the pressure of high rates of Duffy negativity in the human host population would have reduced the population size of *P. vivax* in Africa, resulting in potential genetic bottlenecking and

the associated reduced diversity. Possibly causing the lowered diversity seen across this continent, especially when compared to the Asian population. The caveat to this is that in all current studies, there are very few genome sequences of *P. vivax* from West and Central Africa – one each from Nigeria and Sao Tomé (Culleton *et al*. 2011), and Mauritania (Mu *et al*. 2005). As such, the continent-wide diversity may be larger than currently assumed.

1. 3.3 Current Understanding of Global Distribution Patterns Pv and Pf

Having introduced some of the disease dynamics, and their differences in Pf and Pv, as well as what is known about their possible ages, this section shortly discusses the differences in their distributions, and what is known of the history and other dynamics behind them.

We already know *Pf* is prevalent across much of Africa, with reduced *Pv* presence due to the Duffy negative adaptation in many African populations. Another note on their general differences in distribution pattern is that *Pf* is restricted to tropical regions. As it develops more slowly during the developmental stages in the mosquito, it needs high temperatures to successfully transmit. In addition, it misses the hypnozoite life stage, which allows *Pv* to bridge between seasons lacking mosquito vectors or seasons that become too cold for the parasite to survive in the mosquito host. *Pv* is present, and has been present much more extensively, across both tropical and temperate zones. Temperate zones historically included coastal areas in the UK, the Netherlands and along parts of Scandinavia. The most northern European latitudes for *Pf*, in the recent past and excluding single introductions from travellers, reach into Spain and Italy. Not much is known about species diversity of either parasite in Europe, especially since both were largely eradicated, and any current ones are likely to have been more recent reintroductions (Rodrigues *et al.* 2017).

In Asia, *Pv* is still present in parts of North and South Korea, Japan, Mongolia and Northern China, including Yunnan (Miao *et al.* 2012). It has distinct genetic diversity, present even within the slowly mutating mitochondrial genome, across sub-continental regions from the Pacific islands of Vanuatu in the Southeast, to Korea and Japan in the Northeast, and Bangladesh in the west (Miao *et al*. 2012, Culleton *et al*. 2011). This diversity will be discussed further later on in the thesis (*see Chapter* 2), and has been discussed in general previously, including its possible underlying driving mechanisms (*see Chapter 1.3.2*). In contrast, Asian diversity in *Pf* is both much less diversified and less differentiated, especially in those genetic regions pertaining to historic dynamics (e.g. Schmedes *et al*. 2019, Tanabe *et al.* 2010, Conway 2003). This lack of variation, which is not only seen in Asia but across the globe, is thought to be linked with a later

(re)introduction of *Pf* into the region, due to a more generally delayed spread across the globe of *Pf* compared to *Pv*, as discussed earlier.

Diversity in African *Pv* is very low, with hardly any showing in previous studies of the mitochondrial genome (Culleton *et al*. 2011). Whether this remains true with an increase in samples from the African continent – without Madagascar, which is the main contributor to the African datasets in current publications – is the question. Due to the low diversity found across its current samples, estimating the age of *Pv* presence in Africa is complex. *Pf* diversity in Africa also has a low diversity, though much higher compared to *Pv*, and generally high compared to the genetic diversity of the species overall (e.g. Schmedes *et al*. 2019, Tanabe *et al.* 2010, Conway 2003). The relatively high diversity of *Pf* in Africa, and the presence of ape-associated malaria lineages in Africa that are genetically most closely related to *Pf*, have made placing the origins of *Pf* in Africa relatively easy from early on (Livingstone 1958, Tishkoff *et al.* 2001, Coluzzi *et al.* 2002, Hartl *et al.* 2002, Hartl 2004, Joy *et al.* 2003, Rich *et al.* 2009), confirmed later by the discovery of *Pf*-associated malaria species in African gorillas (Liu *et al.* 2010, Rayner *et al.* 2011). Even so, it must be noted that there is low to no differentiation of the *Pf* mitochondrial genome at a more localised geographic level, across the continent.

For South America, the introduction route and timing for *Pf* have always been clearly linked with the introduction of slaves from Western Africa around the 1600s (Yalcindag *et al*. 2012, Rodrigues *et al*. 2017). Only recently, similar routes and timing for the introduction of *Pv* have been suggested, this time from Southern European origins, coming with the European explorers and traders instead of the West Africans (Rodrigues *et al*. 2017). Though evidence for this theory is mounting, it still remains questionable as it is mainly theoretical, with some support from limited genetics. As a consequence, the route of *Pv* introduction into the Americas remains a little unclear.

The strongest differentiation found in *Pf* mtDNA by the study by Schmedes *et al*. (2019), is the clustering of samples from the Philippines, Southeast Asia, Papua New Guinea and Vanuatu. As with *Pv*, this region appears both more strongly differentiated and more strongly diversified than most other regions. It could be an indication that there are strong selective factors at play across this region, and factors that impact both, regardless of age of introduction – be it climate, geography, human populations or vector species.

The focus of this project is on *Pv* as it is widespread across the Indo-Pacific region and believed to have been present there for a considerably long period of time (Groube 1993, Mu *et al.* 2005), possibly from the first known human migrations, estimated to be over 40k years ago (kya). It is historically the most prolific malaria species, with widespread human genetic adaptations, such as several versions of the

glucose-6-phosphate dehydrogenase (G-6-PD) deficiency gene spread across the globe (Mehta *et al*. 2000, Howes *et al.* 2013). The presence of another malaria-associated mutation, which produces Southeast Asian Ovalocytosis (SAO), found across the Austronesian population, indicates exposure and adaptation to malaria before migration, and the possibility of these human populations forming a source of proliferation (Rosanas-Urgell *et al.* 2012). Though there are many arguments against the ability of *Pv* to cross the Pacific, given time constraints and an absence of *Anopheles* spp. vectors for its further propagation, it is not an unsuitable migrant. Its ability to remain latently present in humans for several years, as so-called hypnozoites, concealed in the liver until circumstances trigger it to re-assert itself in the bloodstream, make *Pv* an opportune migratory disease and give it the ability to survive even in short-term absence of mosquitoes. As immunity to *Pv* is built up relatively fast (Mueller *et al*. 2009) it is generally less debilitating than *Pf*, and those infected with an active form can usually still travel and survive.

The environment across most of ISEA and Melanesia would have been suitable for a pre-Austronesian introduction of malaria. The Papua New Guinea highlands have been estimated to have developed forms of gardening from around 10,000 ya (Denham *et al.* 2003, 2008). Though the highlands themselves would have likely been unsuitable for malaria, such practices present in the coastal areas would have provided additional facilitation for mosquitoes and the spread of malaria well before the Austronesian expansion. The main malaria vector species in Melanesia, which includes members of the *An. punctulatus* complex, are thought to have evolved well before the arrival, and evolution, of the first humans (Logue *et al*. 2013, 2015). The tropical environment and prior presence of *Anopheles* mosquito species would likely have been sufficient for malaria to establish itself. Unlike *Pf*, *Pv* has been indicated to survive amongst small, isolated, migratory indigenous populations in Southern America (Ivo Mueller *pers. comm.* 2013), confirming its suitability as an early migratory disease.

1. 4. Large-Scale Diversity Patterns

1. 4.1 Disease genetics and human migrations

In recent years, there has been a global focus on the spread of disease through human trade, travel and other migratory activities. Humans represent an extremely widespread and well-connected species, with many migrants moving between populations. Numerous studies have recognised that these vast connection networks have had and could have serious implications for the spread of unwanted species, including diseases (Brockmann *et al*. 2006, Tatem *et al.* 2006, Tatem & Hay 2007, Colizza & Vespignani 2010, Bos *et al.* 2011, Chen & Vasilakis 2011). Species propagated through human movement, besides agricultural species for trade and subsistence, and those such as pets that are deliberately exchanged, include: commensals such as rats, mice and house geckoes; and parasites, diseases and disease vectors such as mosquitoes, ticks, yellow fever and hepatitis C (Moritz *et al.* 1993, Austin 1999, Duplantier *et al*. 2002, Matisoo-Smith & Robins 2004, Bryant *et al.* 2007, Searle *et al*. 2009, Hardouin *et al*. 2010, Delatte 2011, Short & Petren 2011, Tonione *et al.* 2011, Markov *et al.* 2012). The spread of some diseases that have migrated with(in) human beings goes back to the very origins of modern humans, such as the bacterium *Helicobacter pylori*, found in the stomachs of around 50% of the current human population (Moodley *et al.* 2012). Furthermore, studying the past of such stowaway diseases also gives an indication of the associated processes of invasion, adaptation and evolution of the disease and occasionally of humans (Lambrinos 2004, Facon *et al*. 2008). This knowledge contributes not only toward understanding the current dynamics behind the diseases studied, but also in predicting the spread patterns of introduced diseases in the future (McNeill 1998, Bryant *et al.* 2007, Markov *et al.* 2012, Moodley *et al.* 2012).

Most recent publications using disease genetics to trace human migration patterns originate from groups focusing on the genetic variation of diseases. A few of them focus on the human past that is likely associated with the observed patterns. One such study of great interest is the previously mentioned study by Moodley *et al*. (2009), which focuses on the ISEA migrations and their place in global human migrations through the DNA of the stomach bacterium *Helicobacter pylori*. Its results are a good indicator for what we expect to find for the phylogeny of *Plasmodium vivax* in a global setting, focusing on the Indo-Pacific migrations. The main discovery was a dichotomy in the genetics of *H. pylori* around ISEA, which is strongly linked to the two different populations found there. The first, associated with Papuans/Melanesians/Non-Austronesians, is linked to the African strains but with a split from other Asian *H. pylori* dating back to around 31,000-37,000 ya, indicating the different Asian strains diverged after humans first emerged from Africa, yet well before the time of the Austronesians migrating into ISEA. They show the presence of different strains across the Asia-Pacific region.

Figure 2 shows the historic movement connecting populations of *H. pylori,* including around the Asia-Pacific area. It shows splits in populations between Asia, East Asia, Papua/Melanesia and Austronesia. Due to the time of infection being restricted to the first weeks of life through mother to child transmission, the genetics of *H. pylori* is a strong proxy for human genetics. This can be seen by the correspondence with human migration estimates from human mitochondrial DNA, as published by Bae *et al.* (2017). Besides demonstrating past origins in the Melanesian area, the study by Moodley *et al.* (2009) goes further, showing broad migratory movements across the globe. The results from Bae *et al.* (2017) also show not all movements have been unidirectional, in contrast with the simplified version as given in *Figure 2*, with influence from populations in eastern Asia diffusing back through to Africa. The further from Africa, the less backflow there appears to be, possibly with oceans blocking/slowing much of the traffic for the Asia-Pacific and American zones.

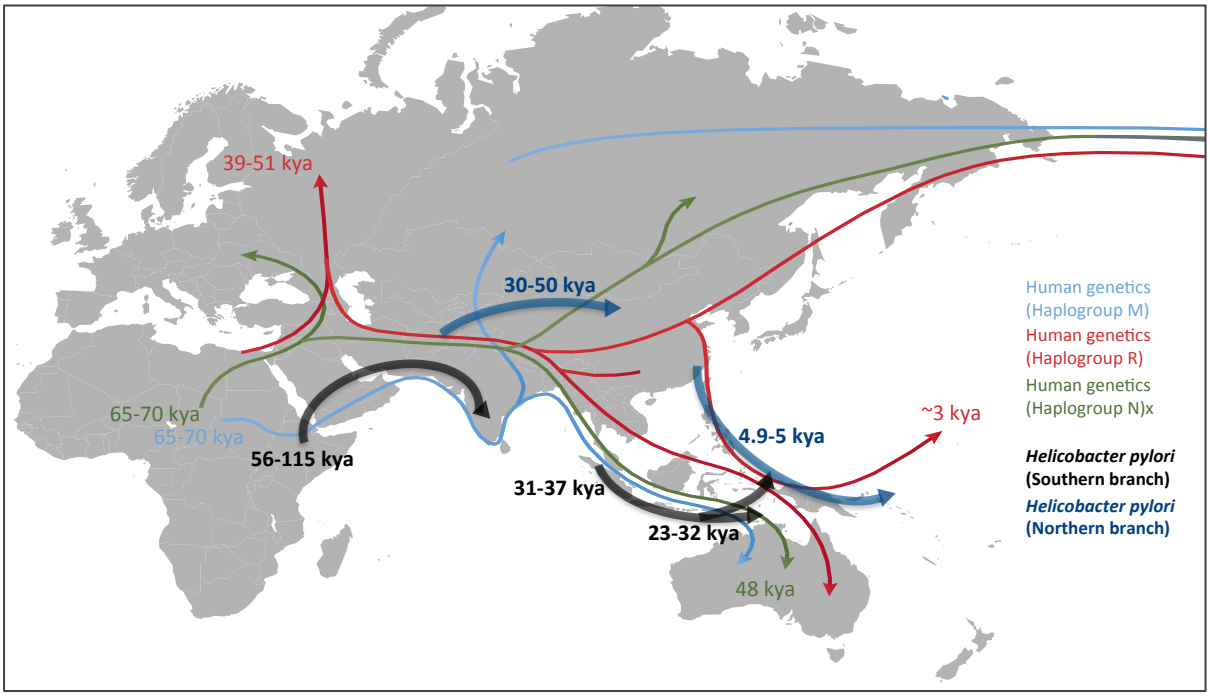

Figure 2: The broader global context of the phylogenetics of *H. pylori* and human migrations based on Moodley *et al.* 2009 and Bae *et al.* 2017. Migration directions of humans according to calculations by isolation with migration models for the genetics of eight different *H. pylori* population pools are given in dark blue and black, denoting the northern and southern routes respectively. Routes as found for the different human mitochondrial genome haplogroups are given in light blue (haplogroup M), red (haplogroup R), and green (haplogroup N). For *H. pylori* dates given are estimated ages of time to the most recent common ancestor of populations across all the regions migrated towards, following an eastward movement. It denotes the estimated earliest time of migrants settling in the region in question.

For *Pv*, a similar framework would be of great interest for understanding human Indo-Pacific migrations, and the resulting genetic variations within a global context. This project aims to understand the patterns of *Pv* distribution and explore the possibilities for using *Pv* in a similar fashion with an extended dataset of *Pv* mitochondrial genomes. An earlier study by Miao *et al.* (2012), involving the

mtDNA of *Pv* and human movements through Asia, indicates a comparable pattern present in *Pv* as seen in *H. pylori*. The paper concentrated on mtDNA diversity, but did not look into the driving forces causing it or estimated the times of introduction. In addition the dataset was relatively small with no or little representation from large geographic regions, including mainland Africa and Asia outside of East Asia and India. These regions are crucial to understand the past dynamics of *Pv* as they cover routes connecting Africa, the supposed region of origin, to East Asia and Melanesia, sites that have high diversity and are well represented in the *Pv* mtDNA dataset.

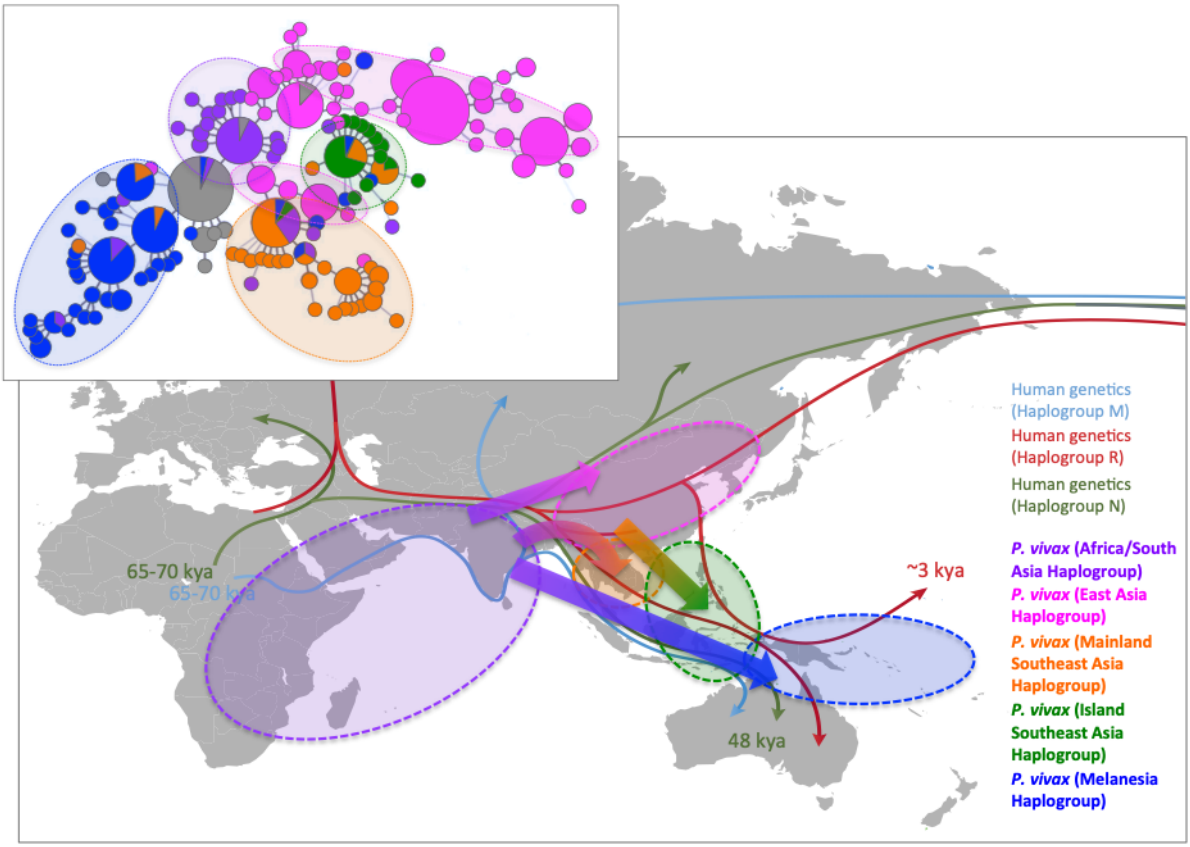

Figure 3: Map representing the main geographic regions for the main haplogroups as given in *Miao et al. 2012: Fig. 2*, and their connections, overlain on a map showing the main human migrations based on data from human mitochondrial genomes, based on that published by Bae et al. 2017. Insert: the haplotype network of *Pv* mt genome sequences by Miao *et al*. 2012, recoloured to match the map. Each circle represents a different haplotype, each colour the region from which a sample originates. The size of the circles indicates the number of samples with that haplotype, the smallest circles indicating a single sample. Each line is a mutational step.

Figure 3 shows a rough geographical representation of the results from the study of mitochondrial DNA (mtDNA) diversity of *Plasmodium vivax* across Asia, and its global connections, published by Miao *et al* (2012). The figure demonstrates a high diversity in the Melanesian region (dark blue), a connection between the Americas and Melanesia (grey and dark blue haplotypes respectively), and connections between African and South Asian samples (purple) that are placed as intermediaries between East Asia, Southeast Asia and Melanesia. The latter connection is explained through migrations not only in the

deep past, but also more recent history, including Indian Ocean trade and later connections through Western colonisation.

Some of these findings overlap with our predictions as set out in the introduction. The differences are that the focus of the paper by Miao *et al*. (2012) is mainly on the genetic diversity found in China and is less concerned with the implications that the migratory dynamics behind the diversity might have for understanding the human past. Even so it remains an extremely useful pilot study for our purposes. Two other papers that provide key insight into the past of *Pv* through mtDNA are by Mu *et al*. (2005) and Culleton *et al*. (2011). Both these papers also show strong association between Melanesian and South/Central American samples, which is unsurprising as both the papers (Culleton *et al.* 2011, Miao *et al*. 2012) incorporate the data from Mu *et al.* (2005) into their study. Two more recent papers focused explicitly on the origin of American *Pv* (Taylor *et al*. 2012, Rodrigues *et al.* 2018). In combination with another paper containing data on an old Spanish sample (Gelabert et al. 2016) these studies have still not satisfactorily resolved the origins of American *Pv*, though there is increasingly strong evidence for an African/European origin, similar to that of *Pf*.

As with *H. pylori* it is possible that the two major human migrations to ISEA and Melanesia could have introduced distinct strains of *Pv.* The connection between *Pv* and humans is not as strong as that of *H. pylori*, making it more complicated to use as a direct proxy for human migrations. However, its greater malignancy means that the implications for its influence on exposed human populations and human genetic evolution are significantly larger, and therefore even a basic understanding of the past between humans and *Pv* is of great interest.

1. 5. Small-Scale Diversity Patterns

1. 5.1 Populations and disease

Besides the global large-scale patterns that have been discussed in the previous section, there is a plethora of smaller scale variation of human genetics and culture found in Melanesia. Current theories on drivers behind population-level diversity contribute towards understanding past interactions between humans and disease, both within the region and in a more general sense. The main focus of this section will be on the influence of disease, malaria in particular, on a population and vice versa. This includes activities by populations that influence disease levels (e.g. through changing their direct environment), as well as the influence of disease on human population dynamics.

Since Groube's (1993) publication on the influence of malaria on human population diversity in Melanesia, there appears to have been only sporadic mention of the theories he puts forward. Though studying the effects of disease on host populations has been an established area of research in the biological sciences for a while (Hamilton *et al.* 1982), Groube's theories have mainly resided within the spheres of archaeology and anthropology (Fincher & Thornhill 2008). Fincher & Thornhill (2008) discuss how infectious disease can act as a wedge between populations, stimulating isolation and thus diversification of populations over time. This has much in common with Groube's theories concerning the population diversification in the Pacific and the role of malaria. Though the theory is interesting enough to be considered, there are several reasons for it failing to have generated wider attention, the reasons for which are laid out in the following paragraphs.

1. 5.2 "Parasite-driven wedge"

Groube's (1993) explanation for much of the human population structure found across the Melanesian region is the burden of malaria. As malaria is a disease that is relatively host density-dependent, small, separate human populations would have been better able to survive malaria, especially as large populations are subject to epidemics (Kermack & McKendrick 1927, 1932, 1933, Bailey 1957, Aaron & May 1982, Bailey & Norman 1982; all in Groube 1993 pg 170, also in Fincher & Thornhill 2008). Groube's theory follows the thought that whilst existing in relative isolation, these populations would build up a balance of immunity and infection with their local malaria strains. Each population would adapt to deal with slightly varying strains of malaria. Any migrants would run the risk of being exposed to a strain of malaria for which they had no immunity. The implication from this would be that populations gained an advantage in isolated existence, beyond contact with specific, close, communities (Groube 1993, Fincher & Thornhill 2008).

Groube calls this a “parasite driven-wedge”, where isolation between populations is increased over time through exposure and adaptation to local parasite strains. If there was movement between populations, this could have happened mainly at the time of marriage. Groube (1993) suggests that as pregnant women would especially run the risk of malaria becoming fatal during pregnancy, it is likely that any marital associated movements would involve residential mobility by the male partner, creating a matrilocal marriage system. Two areas of highest malaria infection in the world, West Africa and Island Melanesia, have a conspicuous predominance of matrilocal marriage in their local cultures, suggesting some possibility of truth in these conjectures.

Fincher & Thornhill (2008) predict that the areas with the highest levels of infections will have the highest levels of population diversity as well as smaller population contact networks, due to the host-parasite co-evolution of highly infected populations inducing isolation, similar to Groube's theories. Their measure of diversity is not genetics but language, a system that is likely to show diversification at an earlier stage, since it is more plastic than genetics. With these predictions in mind, it is interesting that the areas mentioned by Groube (1993) as having high malarial load, are also regions with extremely high cultural diversity. Island Melanesia is an especially strong case for this argument.

From the above it follows that some of the cultural diversity of the Melanesian region could be explained through disease limiting population densities. However, it is wise to approach these theories with caution. Research into strain specific immunity, and the relationship between pregnancy and malaria infection, actually shows a multi-faceted picture that is not as straightforward as what is assumed to come to the conclusions made by the researchers above.

Although there is evidence of local immunity to *Pf* in human populations in Melanesia (Alyssa Barry *pers. comm.* 2018) the argument needs to be approached with care. Besides strain specific immunity to *Pv* there is a build-up of strain transcending immunity (Doolan *et al*. 2009, Mueller *et al*. 2013). In lab tests, mainly for treating syphilis, patients were exposed to only a single strain of *Pv*. Subsequent exposure to different strains would make the patient very ill, though with a slight reduction of especially medical complications (Mueller *et al*. 2013, Snounou & Pérignon 2013, Doolan *et al*. 2009, Langhorne *et al*. 2008). These studies probably led to the early impression that immunity was strongly strain specific. However, once a patient has been exposed to several different strains, as is the case generally in natural populations, the strain transcending immunity becomes more pronounced, significantly reducing illness from exposure to most – if not all – strains within a species (Doolan et al. 2009). For *Pv*, the reduction in severe illness in humans built up from strain-transcending immunity can last up to 3.5 years (Mueller *et al*. 2013) and protection against clinical disease up to 30 years (Langhorne *et al*. 2008).

In regions with continuously high malarial infection, there would therefore be little extra risk in migrating between populations. Even so, the isolation of human populations in Melanesia whilst exposed to high levels of malaria has probably led to the marked diversity in genetic adaptations found in human populations, including a marked variety in G-6-PD (Howes *et al*. 2013) and human leucocyte antigens (HLA, Vina *et al*. 2012, Skoglund *et al*. 2016), and varying prevalence of SAO, Duffy negativity and Gerbich negativity across different Melanesian regions (Zimmerman *et al*. 1999, Patel *et al*. 2004, Rosanas-Urgell *et al*. 2012). For these adaptations to become prevalent would have taken several human generations (Hedrick 2011), and could have influenced genetic structuring in malaria populations too. However, this genetic structuring is expected to be present in other regions of the malaria genome and not the mitochondria. With its slow mutation rate, extremely long-term isolation would be necessary for mutations on the mtDNA to become saturated in a population, besides the presence of only three genes on the mtDNA that could be targets of selection pressure. Some patterns of connection and isolation of human populations on a shorter time-scale have been detected with microsatellites (Koepfli et al. 2015, Fola *et al*. 2018).

The argument for matrilocality due to malaria also has little strength. In populations that are continuously exposed to malaria, the main adults at risk of dying from the disease are women in their first or second pregnancy. Research has moved on since the past assumptions that it was due to the natural immuno-suppression occurring in pregnant women, or that the placenta provides non-immunised shelter for the parasites (Fried *et al.* 1998). Amongst others, to find that though the placenta provides an extra organ for parasite sequestration in *Pf* infections (Beeson & Duffy 2005, Cowman *et al.* 2016), it is not necessarily non-immunised. The placenta itself increases sequestration opportunities for *Pf* infected mature cells, thereby facilitating increased disease. In addition, pregnancy also selects for parasites with a different binding mechanism for invading red blood cells, which also have different antigens expressed on the surface of infected erythrocytes (Rogerson *et al*. 2007, Cowman *et al*. 2016). Due to this, parasites and parasitised cells are generally unrecognised by the immune system, even if the patients have built up immunity to the normal *Pf* parasites from regular exposure (Beeson & Duffy 2005, Rogerson *et al*. 2007, Doolan *et al*. 2009, Cowman *et al.* 2016). Over subsequent pregnancies, the immune system appears to build up recognition of these parasites. According to this theory, all women – local or migrant – are at risk of increased illness during pregnancy from *Pf*. This means there is no perceivable extra benefit to a matrilocal-based society for areas with *Pf* infection.

The risk for increased illness seems mostly absent in *Pv* (Desai *et al*. 2007), or at the very least much reduced compared to *Pf* (Nosten *et al.* 1999) – possibly due to the lack of organ sequestration behaviour. In contrast, increased levels of maternal anaemia and lower the birth weight in the new-born remain present, for both *Pv* and *Pf* (Desai *et al*. 2007, McGready *et al*. 2012, Doolan *et al*. 2009). *Pf* can also lead to

stillbirths and spontaneous abortions, but this feature is argued over in *Pv*. Desai *et al*. (2007) claim no influence, however the data provided and studied by McGready *et al*. (2012) indicates similar rates of abortion for *Pf* and *Pv*. The difference here may be epidemiology. In women who were exposed to *Pv* from early childhood, vivax malaria does not seem to cause much morbidity in pregnancy. However, in non-immune women experiencing their first vivax malaria episode, the morbidity generally is higher in pregnant compared to non-pregnant women. McGready *et al*.'s population is more of the latter, non-immune, type. Even so, according to both *Pv* retains the negative influence on maternal anaemia and birthweight in multigravid women, more so than *Pf*. In short, though there is morbidity associated with infection, for *Pv* it does not appear to be much more significant for non-pregnant women. For *Pf,* there is increased risk, but, as far as is known, this is risk is not dependant on strain, and disease presents even in immune women during their 1st and 2nd pregnancies. Leading to the conclusion that neither disease is likely to play a significant role in separating populations based on the influence it has on pregnant women.

From this, one may conclude that small-scale population separation caused by malaria only is unlikely. Even so, if the effects of malaria were combined with other factors, such as environmental influences, malaria itself could exacerbate already existing barriers between populations. This is believed to be the case in Papua New Guinea (PNG) between the highland and coastal human populations, where the strong separation is probably at least partially due to disease in general and malaria specifically. The area between an altitude of 500 and 1200m is plagued by unstable outbreaks of epidemic malaria. Other environmental factors contribute to the region being seen as a 'population sink' that cannot sustain human life without migrants from other areas (Michael Bangs *pers. comm.* 2012, Ivo Mueller *pers. comm.* 2013). Environmental factors complicating life in the area include high rainfall and steep terrain, complicating agricultural activities, but also disease pressure from both the previously mentioned epidemic malaria and the highland associated pneumonia. Thus this area forms a large and consistent barrier between the highland and coastal areas. It is not unlikely that this could have acted as a wedge, with populations on both sides adapting separately to widely different environments.

Purely environmental influences, such as the island geography of ISEA and Melanesia will have contributed to the diversity in the area, be it on a smaller scale. Geographic separation of populations can result in diversification through lack of continued contact. In biology this is termed the process of allopatric speciation, if it were to result in the populations diversifying into separate species. In the case of humans and culture, a term such as 'allopatric diversification' might be appropriate. The separate development of human cultures through geographic separation by oceanic barriers, however, does not completely explain the patterning of diversity found. The populations are known to have had the ability to cross oceanic barriers, and did so with purpose. It would be unlikely that they settled the Australasian

region if this was not the case, and genetic diversity would have been much lower if settlement had been through accidental crossings (Chappell 1993, Kirch 2000). The earliest Australasian settlers are estimated to have arrived in ISEA from Africa through a Southern Asian route from 60,000 ya to at least 40,000 ya. Even with lowered sea-levels at that time the first human settlers would have had to cross broad, deep sea-channels from Sunda, the Afro-Eurasian continent, to reach the enlarged continent of Sahul, which included a connected New Guinea and Australia (Chappell 1993, Kirch 2000).

Even if discounting the oceanic barriers, the ecology of the area is expected to have played a significant role in population diversification. On the large island of New Guinea, linguistic diversity is one of the most prolific in the world (Spriggs 1997, Kirch 2000, Pawley *et al*. 2005). The geography is varied, including barriers in the form of high mountains, large rivers and mangroves, and both geography as well as ecological variation have been shown to influence cultural and genetic diversity in the region (Stoneking *et al*. 1989, Friedlaender *et al*. 2008, Romaine 2013). Population variation is also increased by the Papuan/Melanesian versus Austronesian descent of the populations, especially as the populations had the opportunity to mix their cultures at different ratios depending on the input of each. With all these factors influencing population diversity, it is possible that the significant presence of disease and its role in selection can explain some of the gaps in understanding the high diversity of the region, but it is hard to tease apart from the multiple dynamics that contributed to it.

1. 5.3 HUMAN ACTIVITIES AND VECTOR-BORNE DISEASES

According to O'Brien and Laland (2012), humans impact their own evolution through influencing their surroundings. This includes the consequences of human activities affecting disease dynamics. In Boyd's book on malaria parasitology (1930), there is the observation that long-distance dispersal of malaria likely occurs through human migrations. In addition, human activities can influence malaria infection rates, and malaria transmission, in several ways. This includes the movement of new susceptible people into an area, increasing infectivity and parasitaemia transmission levels, which in turn exceeds the usual threshold of the local population and creates an epidemic. Alternatively the proliferation of vectors can occur through activities such as land development that increases the number of (stagnant) pools or sunlit pools, for example by obstructing natural drainage or cutting down trees.

One of the early migrations out of Africa is thought to have been connected to early agricultural land clearing (Salamini *et al.* 2002). The development of early agriculture is connected to large-scale population increase, which ultimately led to mass migrations (Cormier 2011). Clearings made in African forests through agricultural practices by humans led to a higher incidence of the mosquito *Anopheles gambiae*,

thus increasing the transmission of *Pf* malaria around 6,000 ya (Livingstone 1958, Tishkoff *et al.* 2001, Coluzzi *et al.* 2002, Hartl *et al.* 2002, Hartl 2004, Joy *et al.* 2003, Rich *et al.* 2009). According to Coluzzi *et al.* (2002) *A. gambiae* also became increasingly anthropophilic, whereas it was previously predominantly a livestock/wildlife feeder. Wilkinson (1994) argues that swidden agriculture, which creates open spaces and allows people to move into forested areas, may have been the main cause of the lowland Mayan civilisation's downfall through an increase of mosquitoes and yellow fever.

Other aspects of agricultural practices, such as stagnant water in cleared areas or flooded agricultural land for crops such as taro and rice, could enlarge the breeding grounds for mosquitoes (Livingstone 1958, Klinkenberg *et al.* 2003, Ohba *et al.* 2012). Alternatively, it has been found that in Southern Asia the more rigorous removal of forests in favour of plantations decreases the occurrence of the malaria-transmitting mosquito *Anopheles dirus* (Obsomer *et al.* 2007), with more dramatic decline of mosquito populations for urban areas. Such contrasting effects of human environmental modification on the increase or decrease of malaria exposure are due to distinct characteristics of particular mosquito species. Some, such as *A. dirus* are typical forest dwellers, whereas *A. sinensis* is commonly found in rice-paddies. An indirect effect of agricultural activities on human disease is that it facilitates for larger populations. The increased density of human populations provides a sustained breeding ground for infectious diseases (McNeill 1998, Oppenheimer 2012). The increase in human vectors within a relatively small area would have been a contributing factor to increasing malaria levels when early land-clearing practices increased mosquito vectors.

Not just agriculture, but other variations in lifestyle among human populations can have a significant effect on the disease presence and manifestation in different populations. This has been shown for Chagas' disease in South America (Coimbra 1988). The disease infects mainly humans in the highlands, but in the lowlands is a zoonotic disease. The difference is likely due to variations in lifestyle influencing host preference of the blood-sucking 'kissing bugs' that are the insect vector of this disease. Such variations include the type of housing, how much the population moves and the presence of domestic animals.

Overall the relationship between human activities, vectors and disease is intricate. ISEA and Melanesia form a good study area to study these relationship due to its environmental diversity and varied cultures, both now and in the past. Though the main *Anopheles* vectors in Melanesia are not strongly anthropophilic, and will readily bite domestic and wild animals, densities of domestic animals are rather low and make humans an accessible target (Ivo Mueller *pers. comm.* 2013).

1. 5.4 HUMAN INVASION, DISEASE AND IMMUNITY

Processes whereby human activities influence mosquito densities and thus disease transmission could have had significant repercussions on populations during the initial stages of such activities, comparable to introductions of novel diseases into previously unexposed populations (McNeill 1998). Newly infected populations often lack the necessary immunity to a well-developed disease, as well as missing any possible cultural adaptations. Impacts of such infections have been demonstrated by the introduction of small-pox on the Native Indian populations of the Americas during European colonisation (Snow & Lanphear 1988, McNeill 1998). In reverse, populations with established diseases could be hard to invade by novel, non-adapted, migrants (McNeill 1998). The latter scenario protects initial settlers, with the disease only facilitating for migrants that are in some way already adapted. This dynamic has been referenced in some cases as the main difficulty in invading and settling malarious regions in Africa and South Asia by Europeans.

Studies on many other species have shown that for species in general it is harder to invade an already established population, especially one that is occupying the available resources, if other things are equal (Stachowicz *et al.* 2002, Hardouin *et al.* 2010). Few studies have looked at what the effect would be if - instead of the invading population - the original population had a natural weapon in the form of disease, in addition to having the advantage of being the initial settlers. The lack of attention for such a 'native over settler' point of view could be due to the more recent experience of Western history being involved with large-scale colonisations. Western colonisations still form a demonstration of this theory, as in contrast to replacing the indigenous populations in Northern America and Australia, they did not manage to settle in significant numbers in the tropical, malarious, regions of Africa and Southern Asia.

The reverse scenario of advantaged early settlers could well be the situation for the Australoid (Papuan) populations colonising Sahul, particularly if they brought malaria with them. The barrier of disease, besides the geographical barrier of the ocean, might provide an explanation for why migrants that populated Asia failed to settle the Melanesian areas to a noticeable degree until the Austronesian influx. A barrier formed by infectious disease could have remained in place until a sea-faring population adapted to the disease arrived and settled the fringe areas. Adapted migrants, especially those bringing in a new strain or species of malaria affecting the earlier settlers, would have less trouble settling in the infected regions (for example *P.vivax* infected and adapted Austronesians, bringing genetic mutations such as SAO with them). As novel exposure to the disease would still affect the settlers, uncolonised areas are likely to have had a preference as migratory destinations.

Groube (1993) discusses an alternative scenario besides the ones mentioned above. As the dynamics of malaria are very much dependent on the population structure and immunity of the host populations, an

imbalance caused by immigrants could affect the colonised populations to such a degree that they are at least as disadvantaged as the migrant populations. The dynamic is as follows: unimmunised migrants enter the area, they get infected and increase the density or prevalence of infection. With a constant presence of migrants, levels of parasitaemia are raised to above the threshold levels of what the local populations are used to, triggering an epidemic that affects migrants as well as locals. This is also known as the 'spillback' effect, as opposed to the 'spillover' effect of new migrants bringing novel diseases with them (Kelly *et al.* 2009). Constant influx of migrants will bring non-infected migrants into the area, whilst most of the local population will be suffering from malaria. There is good reasoning behind this process described by Groube, but it is difficult to see how the migrants would have the advantage. Any new migrants moving into an area exposed to holoendemic malaria, let alone a raging epidemic, are unlikely to avoid being infected, which removes any advantage. At best the playing field would be level. Advantage is given to those with a more rapid development of immunity to the disease, i.e. those that already have an existing level of immunity, or those that have experience of how to deal with the disease. Behavioural adaptation to prior exposure could include knowledge of local cures, if not an already adapted way of life. An example of how rapidly immunity is acquired by non-exposed migrants moving into a holoendemic malarious region is given in the study by Baird *et al*. (2003). Javanese migrants moved into West Papua, suffering severe malarial infections at first exposure, comparable to that of indigenous children. These migrants rapidly gained immunity, but around 7.5% of the infected migrants had to be evacuated due to the severity of the disease, indicating that non-exposed migrants would have a severe disadvantage when entering a malarious region.

1. 5.5 MALARIA AND HUMAN GENETIC ADAPTATION

Besides high diversity in general human genetics, the region of focus in this study is also rich in genetic mutations related to malaria adaptations (Serjeantson *et al.* 1992, Howes *et al.* 2013). Several adaptations are exclusive to the Melanesian area, or local even within certain areas of Melanesia, suggesting high levels of population isolation. However, there is evidence of enough contact between populations to spread the most effective genetic adaptations against the disease (Serjeantson *et al.* 1992). The ages of several human genetic adaptations to malaria have been estimated by different studies; an overview of this is given in *Table 1* (Hedrick 2011). These date estimates vary depending on the study, due to differences in assumptions made regarding the basic values used in the calculations, such as the chances of survival in malarious areas of individuals with adaptive mutations versus those without (the selection coefficient) or reproductive rates of those that survive (Hedrick 2011). As demonstrated in *Table 1*, small variations in these assumptions can give substantially different results. Additionally there is substantial variation in the standard deviation estimates within studies. More rigorous methods would be useful and currently anything based too heavily on precise dates taken from these calibrations would seem questionable.

However, even with the variability of the estimates of times of origin it remains clear from most data in the table that populations in ISEA have long been exposed to malaria infections.

There is evidence for a link between malaria and Austronesian migrants through the presence of the G-6-PD deficiency gene in Vanuatu populations (Kaneko *et al.* 1998). The frequency of the G-6-PD deficiency variant (allele) in populations across the different islands and regions of Vanuatu corresponds strongly to the level of malarial infection present in each area. The areas with a high malarial load contain human populations with a higher frequency of the G-6-PD deficiency gene than those areas with low or no malaria presence. All areas, including those that have had little to no gene flow between them, contain the same G-6-PD deficiency mutation. This suggests the humans came into the area carrying the variant, and those populations still carrying it have been continuously exposed to malaria (Kaneko *et al.* 1998).

Table 1: Overview of the estimated age of alleles in generations and years (for a 25-year generation length) that confer resistance to malaria. *Table after Hedrick 2011, pg. 289: Table 2.* Selection coefficients with an * are assumed values, not estimated, and — indicates no selection (neutrality) is assumed.

Gene allele	Age		Selection coefficient (s)	Reference
	Generations	Years		
b-globin				
S	45–70	1125–1750	0.152*	Currat et al. (2002)
	10–28	250–700	—	Modiano et al. (2008)
C	75–150	1875–3750	0.04–0.09	Wood et al. (2005)
	38–120	950–3000	—	Modiano et al. (2008)
E	100.3 (62–222)	2508 (1550–5550)	0.079 (0.035–0.099)	Ohashi et al. (2004)
G6PD				
A-	254.3	6357 (3840–11 760)	0.044	Tishkoff et al. (2001)
A-	93	2325 (1200–3862)	—	Sabeti et al. (2002)
A-	1800	45 000 (25 000–65 000)	—	Verrelli et al. (2002)
A-	40	1000	0.25	Slatkin (2008)
Med	133.2	3330 (1600–6640)	0.034	Tishkoff et al. (2001)
Med	400	10 000 (0–35 000)	—	Verrelli et al. (2002)
Mahidol	63	1575	0.23	Louicharoen et al. (2009)
Duffy				
ES (null)	1323 310 490	33 075 (6500–97 200)	—	Hamblin and Di Rienzo (2000)
ES (null)	310	7750 (3625–13 125)	—	Seixas et al. (2002)
ES (null)	490	12 250 (4250–26 500)	—	Seixas et al. (2002)
ABO				
O01	—	1.15 million	—	Calafell et al. (2008)
O02	—	2.5 million	—	Calafell et al. (2008)
HLA-B				
B53	100	2500	0.028*	Hill et al. (1991)
B53	86	2150	0.041*	Hedrick 2011

Other evidence for a link between malaria and a specific group of migrants in the Asia-Pacific region is the occurrence of a *Pv* resistance variant associated with Southeast Asian Ovalocytosis (SAO), in certain ethnic groups (Mgone *et al.* 1996, Patel *et al*. 2004). The mutation mainly protects against *Pv* infections, yet in addition it appears to help reduce the severity of *Pf* induced cerebral malaria (Genton *et al.* 1995, Rosanas-Urgell 2012). The SAO mutation is lethal in homozygous individuals (Genton *et al.* 1995), which argues that to remain present in the population there must have been a high incidence of *Pv* malaria. Through having a long-term selective advantage in the heterozygous individual, such mutations with negative consequences will be maintained in the population (Livingstone 1958, 1964).

A possible consequence of contact between two populations carrying different adaptive variants with adverse health effects is that the mutations can accumulate in offspring with mixed ancestry. Instead of conferring an advantage, the decreased fitness of mixed offspring could create a barrier to population admixture, besides any cultural barriers, as mixed progeny would have lower chances of survival to an extent that outweighs any relative advantage gained over malaria due to accumulation of deleterious mutations. Where separated populations adapted to malaria in different ways such a risk would be an added barrier.

1. 6. CONCLUSION

This chapter has demonstrated that the human cultural and genetic diversity found across the Eastern Hemisphere, and especially within ISEA and Melanesia, is complex and includes a diverse array of contributing factors. Whilst substantial research has been conducted on the topic of *Pv* mitochondrial diversity, there is still much left unexplained and there have been no comprehensive studies looking into several influencing factors at the same time. Much of the past dynamics behind the distribution of *Pv* and its connections with the human past has been left unexplored. Though the complexity and scale of the area tackled in this thesis are considerable, it is precisely this diversity that will make for interesting comparisons and help gain a more thorough understanding of factors impacting the interactions between humans and malarial disease. In addition it will aid in identifying avenues for future research into parasite and human evolution combined.

To such an end this project aims to augment a thorough comprehension of the dynamics behind the diversity of *Pv* in the Eastern Hemisphere. There are several factors that have been shown to influence genetic variation on smaller geographic scales or are expected to have a role in the distribution of *Pv*. *Chapter 2* explores the distribution and history of, and interaction between, the two vectors on a global scale. Studying each vector separately without accounting for the influence of the other could overlook important interactions or lead to misinterpretation of data patterns. The expectation is that *Pv* distribution follows human migration and settlement patterns, possibly dating back to the emergence of *Homo sapiens* from Africa, though the emergence could have occurred earlier, with other primates, or later. As strong influence of mosquito vector species on the *Pv* genome has been found in recent publications, the expectation is that patterns in human population distribution could be influenced by alternative underlying patterns in anophelene species.

Chapter 3 takes a closer look at the influence of past human migration dynamics on the mtDNA variety of Pv within a smaller regional context. It looks at the origin of *Pv* in Melanesia, set within the context of the human past both locally and in the surrounding regions. As previous studies show a high diversity of *Pv* in this region there are two alternative, yet compatible, hypotheses. Firstly, the introduction of *Pv* into Melanesia occurred very early on in time, for example with the early *H. sapiens* Papuan populations, and as such it has had ample time to diversify. Alternatively, or additionally, there were several introductions of *Pv*, possibly from different origins such as from both Papuan and Austronesian populations. Attempts at studying smaller scale dynamics in this region failed due to the low mutation rate of mtDNA being unable to pick up on more recent historical dynamics.

The functionality and remaining use for the mitochondrial genome for *Pv*, and malaria in general, with the introduction of whole genome sequencing is the focus for *Chapter 4*. The chapter reviews the use of different genetic sequences for varying research purposes in the past, with a strong emphasis on research in malaria phylogenetics and population structure. As whole genome sequencing is still relatively slow and costly - both resource and time wise - the expectation is that the mitochondrial genome, and its large publicly available database, will still remain a useful tool. In addition, the knowledge garnered on the genome throughout the past years, including an understanding of its relatively consistent and stable mutation rate across the genome, is anticipated to retain its value especially when looking into the past of *Pv*. The chapter also explores the possibility of combining mtDNA data with data from microsatellites as combining data from mtDNA with that of other parts of the genome could increase resolution in phylogenetics and population dynamics, leading to greater insight over slightly smaller time periods than that covered by the mitochondrial genome.

Overall this thesis attempts to answer how *Pv* and its diversity and distribution are connected to the human past. It assesses the value of using the mitochondrial genome as a tool to answer this question and explores what and how other dynamics influence the connections seen. Besides providing an increased understanding of the past of *Pv*, comprehending the influence of these dynamics could provide helpful insights for future research into the past of humans, their migration dynamics and their adaptation to the disease.